

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/836,075DATE: 10/08/97
TIME: 13:05:11

INPUT SET: S20848.raw

This Raw Listing contains the General
Information Section and those Sequences
containing ERRORS.

Does Not Comply
Corrected Diskette Needed

SEQUENCE LISTING

- 1
2
3 (1) General Information:
4
5 (i) APPLICANT: MAERTENS, GEERT
6 STUYVER, LIEVEN
7
8 (ii) TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
9 AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
10 AGENTS
11
12 (iii) NUMBER OF SEQUENCES: 207
13
14 (iv) CORRESPONDENCE ADDRESS:
15 (A) ADDRESSEE: ARNOLD, WHITE & DURKEE
16 (B) STREET: P.O. BOX 4433
17 (C) CITY: HOUSTON
18 (D) STATE: TEXAS
19 (E) COUNTRY: USA
20 (F) ZIP: 77210-4433
21
22 (v) COMPUTER READABLE FORM:
23 (A) MEDIUM TYPE: Floppy disk
24 (B) COMPUTER: IBM PC compatible
25 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
26 (D) SOFTWARE: Microsoft Word 6.0 / ASCII text output
27
28 (vi) CURRENT APPLICATION DATA:
29 (A) APPLICATION NUMBER: 08/836,075
30 (B) FILING DATE: 21 Apr 1997
31
32 (vii) PRIOR APPLICATION DATA:
33 (A) APPLICATION NUMBER: PCT/EP95/04155
34 (B) FILING DATE: 23 Oct 1995
35
36 (vii) PRIOR APPLICATION DATA:
37 (A) APPLICATION NUMBER: EP 94870166.9
38 (B) FILING DATE: 21 Oct 1994
39
40 (viii) PRIOR APPLICATION DATA:
41 (A) APPLICATION NUMBER: EP 95870076.7
42 (B) FILING DATE: 28 Jun 1995
43
44 (ix) ATTORNEY/AGENT INFORMATION:
45 (A) NAME: KAMMERER, PATRICIA A.

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/836,075DATE: 10/08/97
TIME: 13:05:13

INPUT SET: S20848.raw

46 (B) REGISTRATION NUMBER: 29,775
47 (C) REFERENCE/DOCKET NUMBER: INNS:004
48

ERRORED SEQUENCES FOLLOW:

893 (2) INFORMATION FOR SEQ ID NO: 22:
894
895 1 (i) SEQUENCE CHARACTERISTICS:
--> 896 (A) LENGTH: 48 amino acids
--> 897 (B) TYPE: amino acid
--> 898 (D) TOPOLOGY: linear
899
900 (ii) MOLECULE TYPE: peptide
901
902
903
--> 904 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:
905
906 Met Ser Thr Asn Pro Lys Pro Gln Arg Lys Thr Lys Arg Asn Thr Asn
907 1 5 10 15
908
909 Arg Arg Pro Gln Asp Val Lys Phe Pro Gly Gly Gly Arg Ser Leu Ala
910 20 25 30
911
912 Glu Tyr Thr Cys Ala Arg Arg Gly Lys Leu Arg Arg Ser Ser Met Gly
913 35 40 45
914
915

950 (2) INFORMATION FOR SEQ ID NO: 24:
951
952 (i) SEQUENCE CHARACTERISTICS:
--> 953 (A) LENGTH: 149 amino acids
954 (B) TYPE: amino acid
955 (D) TOPOLOGY: linear
956
957 (ii) MOLECULE TYPE: peptide
958
959
960
961 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:
962
963 Asp Gly Ile Asn Tyr Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser
964 1 5 10 15
965
966 Ile Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Val Pro Ala Ser Ala
967 20 25 30
968
969 Val Gln Val Lys Asn Thr Ser His Ser Tyr Met Val Thr Asn Asp Cys

All next page

RAW SEQUENCE LISTING PATENT APPLICATION US/08/836,075

DATE: 10/08/97
TIME: 13:05:15

INPUT SET: S20848.raw

```

970          35          40          45
971
972      Ser Asn Ser Ser Ile Val Trp Gln Leu Lys Asp Ala Val Leu His Val
973          50          55          60
974
975      Pro Gly Cys Val Pro Cys Glu Arg His Gln Asn Gln Ser Arg Cys Trp
976      65          70          75          80
977
978      Ile Pro Val Thr Pro Asn Val Ala Val Ser Gln Pro Gly Ala Leu Thr
979          85          90          95
980
--> 981      Arg Gly Leu Arg Thr His Ile Asp Thr Ile Val Ala er Ala Thr Val 1
982          100          105          110
983
984      Cys Ser Ala Leu Tyr Val Gly Asp Phe Cys Gly Ala Val Met Leu Val
985          115          120          125
986
987      Ser Gln Phe Phe Met Ile Ser Pro Gln His His Ile Phe Val Gln Asp
988          130          135          140
989
990      Cys Asn Cys Ser Ile
991      145
992

```

```

1161      (2) INFORMATION FOR SEQ ID NO: 30:
1162
1163      (i) SEQUENCE CHARACTERISTICS:
--> 1164          (A) LENGTH: 149 amino acids
1165          (B) TYPE: amino acid
1166          (D) TOPOLOGY: linear
1167
1168      (ii) MOLECULE TYPE: peptide
1169
1170
1171
1172      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:
1173
1174      Asp Gly Ile Asn Phe Ala Thr Gly Asn Leu Pro Gly Cys Ser Phe Ser
1175      1          5          10          15
1176
1177      Ile Phe Leu Leu Ala Leu Leu Ser Cys Leu Thr Val Pro Ala Ser Ala
1178          20          25          30
1179
1180      Ile Asn Tyr Arg Asn Val Ser Gly Ile Tyr Tyr Val Thr Asn Asp Cys
1181          35          40          45
1182
1183      Pro Asn Ser Ser Ile Val Tyr Glu Ala Asp His His Ile Leu His Leu
1184          50          55          60
1185
1186      Pro Gly Cys Val Pro Cys Val Arg Glu Gly Asn Gln Ser Arg Cys Trp
1187      65          70          75          80
1188
1189      Val Ala Leu Thr Pro Thr Val Ala Ala Pro Tyr Ile Gly Ala Pro Leu

```

See next page

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/836,075DATE: 10/08/97
TIME: 13:05:17

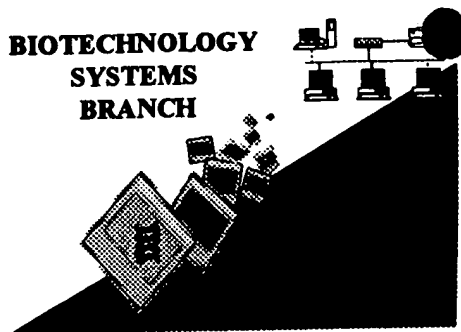
INPUT SET: S20848.raw

1190 85 90 95
1191
1192 Glu Ser Leu Arg Ser His Val Asp Leu Met Val Gly Ala Ala Thr Val
1193 100 105 110
1194
--> 1195 Cys Ser Ala Leu Tyr Ile Gly Asp Xaa Cys Xa Gly Leu Phe Leu Val
1196 115 120 125
1197
1198 Gly Gln Met Phe Ser Phe Arg Pro Arg Arg His Trp Thr Thr Gln Asp
1199 130 135 140
1200
1201 Cys Asn Cys Ser Ile
1202 145
1203

1204 (2) INFORMATION FOR SEQ ID NO: 31:
1205
1206 (i) SEQUENCE CHARACTERISTICS:
--> 1207 (A) LENGTH: 447 base pairs → 448 shown
1208 (B) TYPE: nucleic acid
1209 (C) STRANDEDNESS: single
1210 (D) TOPOLOGY: linear
1211
1212 (ii) MOLECULE TYPE: cDNA
1213
1214 (iii) HYPOTHETICAL: NO
1215
1216 (iii) ANTI-SENSE: NO
1217
1218
1219
1220 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:
1221
1222 GACGGGATCA ATTATGCAAC AGGGAACCTT CCCGGTTGCT CTTTTTCTAT CTCCTCTTG 60
1223
1224 GCACTCCTCT CGTGCCCTGAC TGTTCCTCGCT TCGGCCATTA ACTACCGCAA CACCTCGGGC 120
1225
1226 ATCTACCACG TCACCAATGA CTGCCCCGAAC TCGAGCATAG TTTATGAGGC CGACCACCAC 180
1227
1228 ATCTTGCACC TTCCAGGTTG CGTGCCCTGC GTGAGAACTG GGAATCAGTC ACGTTGCTGG 240
1229
1230 GTGGCCCTTA CTCCTACCGT CGCAGCGCCA TACATCGGCG CACCGCTTGA GTCTCTGCGG 300
1231
1232 AGTCATGTGG ATCTGATGGT GGGGGCTGCC ACTGTTTGCT CAGCCCTTTA CATCGGGGAT 360
1233
--> 1234 TTGTGTGGCG GCTTGTCTTT GGTGGGTGAG ATGTTTTGTT TCGGACCACCG ACGCCACTGG 420 421
1235 group of 11
1236 ACTGCCCAGG ATTGCAATTG TTCTATC 447 448
1237

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION US/08/836,075DATE: 10/08/97
TIME: 13:05:21**INPUT SET: S20848.raw**

Line	Error	Original Text
866	Entered (310) and Calc. Seq. Length (0) differ	(A) LENGTH: 310 base pairs
896	Unknown or Misplaced Identifier	(A) LENGTH: 48 amino acids
897	Unknown or Misplaced Identifier	(B) TYPE: amino acid
898	Unknown or Misplaced Identifier	(D) TOPOLOGY: linear
904	Wrong Or Missing Strandedness Value	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:
904	Wrong or Missing Sequence Topology	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:
953	Entered (149) and Calc. Seq. Length (148) differ	(A) LENGTH: 149 amino acids
981	Wrong Amino Acid Designator	Arg Gly Leu Arg Thr His Ile Asp Thr Ile Val Ala er Ala Th
981	Wrong Amino Acid Designator	Arg Gly Leu Arg Thr His Ile Asp Thr Ile Val Ala er Ala Th
1164	Entered (149) and Calc. Seq. Length (148) differ	(A) LENGTH: 149 amino acids
1195	Wrong Amino Acid Designator	Cys Ser Ala Leu Tyr Ile Gly Asp Xaa Cys Xa Gly Leu Phe
1207	Entered (447) and Calc. Seq. Length (448) differ	(A) LENGTH: 447 base pairs
1234	# of Sequences for line conflicts w/ running total	TTGTGTGGCG GCTTGTTCTT GGTTGGTCAG ATGTTT



Notice of Availability of Checker Program

Applicant Aid for Biotechnology Computer Readable Form (CRF) Sequence Listing Submissions

The Patent and Trademark Office (PTO) has developed a computer program, called Checker, that will aid applicants in identifying and correcting errors prior to making submissions for compliance with the Requirements for Patent Applications Containing Nucleotide Sequence and/or Amino Acid Sequence Disclosures (Sequence Rules: 37CFR 1.821 through 1.825). Final rules were published in the *Federal Register* (55 FR18230) on May 1, 1990, and in the PTO *Official Gazette* (1114 Off.Gaz.PatOffice 29) on May 15, 1990.

Checker is a DOS-based software program that is intended to assist users in determining whether errors may be present in the sequence listings, and is not intended to guarantee that the submission is error-free.

The most current version of the software is available via computer downloading, details are below. Copies on diskette are also available. Updated software versions will not be automatically mailed out; any updates will be announced in the PTO *Official Gazette*.

The software can be accessed/requested from the following locations:

- 1) Dial-up access through the Internet. Location is <ftp://ftp.uspto.gov>
The software is in current directory: [pub/checker/](ftp://ftp.uspto.gov/pub/checker/)
Download all the files. Cost: Free-of-charge
- 3) For diskette copies, mail to: U.S.P.T.O., OEIP, CRYSTAL PARK 3, SUITE 441
WASHINGTON DC 20231

COST FOR DISKETTE IS \$ 25.00

METHOD OF PAYMENT:

Check payable to Commissioner of Patents and Trademarks
VISA/ Mastercard/ Charge- Charges can be faxed to 703-306-2737
PTO Deposit Account

*For Further Information, Contact: **Arti Shah at 703-308-4212***